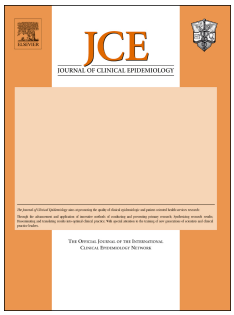
Multivariate meta-analysis helps examine the impact of outcome reporting bias in Cochrane rheumatoid arthritis reviews

Frosi, Giacomo; Riley, Richard D.; Williamson, Paula R.; Kirkham, Jamie J.

DOI:
10.1016/j.jclinepi.2014.11.017

Accepted Manuscript



Multivariate meta-analysis helps examine the impact of Outcome Reporting Bias in Cochrane rheumatoid arthritis reviews

Giacomo Frosi , Richard D. Riley , Paula R. Williamson , Jamie J. Kirkham

PII: S0895-4356(14)00495-8

DOI: 10.1016/j.jclinepi.2014.11.017

Reference: JCE 8760

To appear in: *Journal of Clinical Epidemiology*

Received Date: 1 May 2014

Revised Date: 24 September 2014

Accepted Date: 24 November 2014

Please cite this article as: Frosi G, Riley RD, Williamson PR, Kirkham JJ, Multivariate meta-analysis helps examine the impact of Outcome Reporting Bias in Cochrane rheumatoid arthritis reviews, *Journal of Clinical Epidemiology* (2014), doi: 10.1016/j.jclinepi.2014.11.017.

Multivariate meta-analysis helps examine the impact of Outcome Reporting Bias in Cochrane rheumatoid arthritis reviews

Giacomo Frosi[1*], Richard D. Riley[2], Paula R. Williamson[1], Jamie J. Kirkham[1]

[1]Department of Biostatistics, University of Liverpool, 1st Floor Duncan Building, Daulby Street, Liverpool, L69 3GA.

[2] School of Health and Population Sciences, and School of Mathematics, University of Birmingham, Edgbaston, Birmingham, B15 2TT

*To whom the correspondence should be addressed: giacomo1@liverpool.ac.uk

Abstract

Objective: Outcome reporting bias (ORB) is a threat to validity of systematic reviews. Multivariate meta-analysis (MVMA) can potentially reduce the impact of ORB when outcomes are correlated. The aim of this study was to assess ORB in Cochrane systematic reviews of rheumatoid arthritis, and to demonstrate how MVMA may examine its impact.

Study Design and Setting: Reviews were assessed for ORB in relation to eight outcomes for rheumatoid arthritis using a nine-point classification system. Impact of ORB was assessed by comparing estimates from univariate and MVMA models.

Results: ORB assessment was applied in 21 included reviews, and all contained missing data on at least one of the eight outcomes. ORB was highly suspected in 247 (22%) of the 1118 evaluable outcomes from 155 assessable trials. MVMA and univariate results sometimes differed importantly. The maximum change in treatment effect estimate between MVMA and univariate approach was found to be 176% for one of the outcome considered.

Conclusions: ORB has the potential to affect the conclusions in meta-analyses. This could be avoided if trialists reported on all measured outcomes in full. If missing outcome data are unobtainable, MVMA is useful to examine the impact of missing outcomes and ORB on conclusions.

What is new?

Key findings

- All 21 reviews considered contained missing data from the included primary studies in at least one outcome. High suspicion outcome reporting bias was suspected in nearly a

1. Introduction

Selective outcome reporting occurs when a subset of the originally recorded outcome variables in a trial are selectively reported in a publication based on their results. When outcome reporting is driven by the significance and/or direction of the effect size (e.g. non-significant outcomes are reported only as p-value>0.05, or are suppressed altogether), we refer to this as Outcome Reporting Bias (ORB) [1]. Empirical evidence has shown that outcomes which are statistically significant are more likely to be fully reported compared to non-significant outcomes [2]. Findings from the ORBIT (Outcome Reporting Bias in Trials) study have shown that over a half of reviews did not include full data for a single review primary outcome of interest from all eligible trials and half of the trials assessed with missing data were under high suspicion of ORB [3]. Related work looking at all published reviews from the Cochrane Cystic Fibrosis group revealed that all of the

eligible trials did not include full data when looking across all review outcomes, i.e. both primary and secondary outcomes [4].

One way to reduce the problem of ORB is the introduction of an agreed minimum set of standardised outcomes, to be measured and reported in all trials for a particular disease or condition, referred to as a 'Core Outcome Set' (COS) [5]. In 1994, the OMERACT (Outcome Measures in Rheumatology) group, aided by meetings of experts from international organisations, ratified one of the first core outcome sets for rheumatoid arthritis [6]: tender joints, swollen joints, pain, physical global assessment, patient global assessment, function and acute phase reactants. In studies lasting at least one year, an additional recommendation was that radiographs of the joints to be taken to assess radiological damage.

An observational review of rheumatology trials published up to 2009 has suggested that between 60-70% of trialists conducting trials in rheumatoid arthritis measured the rheumatoid arthritis core outcomes [7]. Encouragingly, 90% of trialists contacted said they would consider measuring the rheumatoid arthritis core outcome set if they were to lead a new trial in rheumatoid arthritis. Nevertheless, knowing that an outcome was measured in a trial does not necessarily constitute appropriate reporting in the trial publication. Another problem in rheumatology trials is that composite outcomes are often reported in place of the individual core outcomes; examples include Disease Activity Score [8] and American College of Rheumatology criteria [9]. The consequence of this type of reporting is that many meta-analyses of the individual core outcomes will contain missing data, when it is known that the individual outcomes were measured and possibly analysed.
If ORB is suspected in a review, and missing data are unobtainable from trial authors, a sensitivity analysis should be considered to determine how robust meta-analysis conclusions are to ORB. Review conclusions that are not robust to ORB may need to be amended because the treatment effect estimates can be overestimated as non-significant outcome data are supressed from the analyses. The maximum bias bound [10]; a multivariate meta-analysis (MVMA) approach [11] and a model-based correction method [1] are three such sensitivity analysis approaches that have been used to assess robustness of the meta-analysis results. The model-based correction method offers a more elegant technique for adjusting for ORB than the maximum bias bound approach as the mechanism for bias is modelled directly [1]. However, both these methods can only be applied to

single outcomes whereas the MVMA approach can be applied to multiple outcomes, which is an optimal approach when considering the simultaneous adjustment of bias in a core outcome set. In this paper we consider only the MVMA approach.

The MVMA approach jointly syntheses multiple correlated outcomes [10], such as the core set of outcomes for rheumatoid arthritis. For example, patients with swollen joints might experience more pain, such that there is a positive correlation between the outcomes swollen joints and pain. It has been shown both analytically [12] and through simulation and application [11] that utilising the correlation allows one to 'borrow strength' across outcomes; in other words, one can learn about unreported outcomes through the reported results for other correlated outcomes. For this reason, MVMA can produce summary treatment effect estimates that are different to those from a traditional 'univariate' meta-analysis of each outcome separately. This is especially important when there are missing outcomes across studies due to ORB [11].

The aims of this current study were to determine the prevalence of ORB in trials included in Cochrane systematic reviews of rheumatoid arthritis by considering all eight outcomes in the core outcome set; and to illustrate the use of the MVMA method. This is the first study to assess ORB in reviews against a core outcome set. We also illustrate the potential benefit of the MVMA approach by applying it to one of the reviews where ORB is likely and the outcomes under consideration have strong correlations. This is particularly important as a previous study which compared univariate meta-analysis with multivariate meta-analysis led the authors to conclude that the choice between univariate and multivariate meta-analysis had unclear or limited practical importance [13].

2. Methods

A cohort of systematic reviews published by the Cochrane Musculoskeletal Group (up to and including the September 2012 issue) that considered pharmacological interventions [disease modifying anti-rheumatic drugs (DMARDs), biologics or glucocorticoids] for the treatment of rheumatoid arthritis were included. Reviews were identified by JJK via the Cochrane topics link (http://onlinelibrary.wiley.com/book/10.1002/14651858/topics) and were those that were indexed under 'Musculoskeletal', 'Rheumatoid Arthritis', 'Treatment [Pharmacological Interventions]' and

'Biologics/Steroids/DMARDs'. Selections were checked by GF. The scope of the COS was not specifically designed for non-drug trials and does not focus on measures of safety; reviews were therefore excluded if they considered non-pharmacological interventions or considered drug safety only. Overviews and reviews that contained no eligible randomised controlled trials (empty reviews) were also excluded, as an assessment of primary studies would not be possible.

All reports for each randomised trial included in eligible reviews were obtained for evaluation. Trials that had been excluded in the 'characteristics of excluded studies' section were also checked for any suggestion of ORB. For example, if a review had excluded trials as a result of 'no relevant outcome data', then these trials were also scrutinised for the presence of ORB and included in the assessment.

2.1 Assessing randomised controlled trials for ORB in systematic reviews

The methodology for assessing ORB in trials follows that of the original ORBIT study [3]. An outcome matrix was constructed using the ORBIT matrix generator **http://ctrc.liv.ac.uk/orbit/** for each review by listing all the eligible studies as rows and the eight core outcomes as columns in the matrix [14]. For each outcome, full, partial and no reporting were distinguished from the information provided in the review. One example matrix for this particular application is given in web appendix 1 (Table 1). All reports for each randomised trial, partially reporting or not reporting on at least one of the core outcomes were obtained and an assessment of the potential risk of ORB was made for each missing or incompletely reported core outcome using the nine-point classification system developed in the ORBIT study [3]. The classification system developed for this study is provided in Table 1. The system identifies whether there is evidence that the outcome was measured and analysed but not reported in full (A to D), whether the outcome was measured but not necessarily analysed (E and F), if it is unclear whether the outcome was measured (G and H), or of it was clear the outcome was not measured (I). If composite outcome criteria were reported in full, but no data on any of the individual core outcomes were reported then the low risk F-classification was used for all core outcomes as it may not have been the trialists intention to analyse the individual core outcomes. If a trialist selectively reported some of the outcomes from the composite, then the high-risk E-classification was used for the core outcomes not reported, as

in this situation it is more likely that all of the core outcomes would have been analysed and the likely reason for some of them not being reported is a non-significant result. In some cases it is often problematic to assess whether an outcome was measured and clinical judgement is required. If there were any uncertainties with the classifications, review authors were consulted. All trials were independently classified by GF and JJK and disagreements were resolved through discussion.

2.2 Assessing the impact of ORB using multivariate meta-analysis

To illustrate the impact ORB may have in a review containing multiple outcomes, a multivariate meta-analysis approach was applied to a review comparing Auranofin with placebo for the treatment of rheumatoid arthritis [15]. The aim was to examine whether the summary results and conclusions from the multivariate meta-analysis differed to those from the original meta-analyses performed by the review authors. The impact was assessed in terms of the change in the treatment effect estimates, change in the statistical significance of the treatment effect estimates and change in the precision of the treatment effect estimates, for each of the outcomes of interest.

First univariate fixed effect meta-analysis (UFMA) and multivariate fixed effect meta-analysis (MFMA) models were fitted, as the original review assumed fixed treatment effects for all outcomes. Secondly, univariate random effects meta-analysis (URMA) and multivariate random-effects meta-analysis (MRMA) models were then fitted for comparison, on the basis that the review commented that in some cases heterogeneity existed, as assessed using I^2 (I^2 ranged from 0% to 83% for the seven core outcomes meta-analysed).

An analogous equivalent to the univariate I^2 statistic which has been proposed for use in multivariate meta-analysis I_R^2 was also computed [12]. Specifications for the bivariate fixed and random effects models are provided in Riley et al. [16] (the interpretation of random effects models is discussed by Higgins et al. [17]); these models were simply extended to the multivariate case [10] as described by Jackson et al. Web appendix 2 provides details of both the univariate and multivariate models in full, with technical details on how within-study and between-study correlation for the multiple outcomes was accounted for. In particular, within-study correlations (of the treatment effect estimates for the outcomes) were derived from the patient-level correlations

amongst the outcomes as provided directly by one of the include trials. The model parameters were estimated using the '*mvmeta'* module [18] in STATA using the method of maximum likelihood for fixed effects models and restricted maximum likelihood method for random-effects models. We did not consider radiological damage in our analyses, firstly because this outcome was not meta-analysed in the original review and secondly as only two of the nine included trials were 52 weeks or longer in duration, meaning that there was no OMERACT recommendation to measure this outcome in the majority of trials included in this review.

3. Results

3.1 Assessment of systematic reviews

3.1.1 Review eligibility

The Cochrane Musculoskeletal Group published 56 unique rheumatoid arthritis reviews up to and including the September 2012 issue (Figure 1). Thirty-five reviews were excluded: 20 focused on non-pharmacological interventions, thirteen studied symptom modifying anti-rheumatic drugs, one was an overview and one focussed on safety. Of the remaining 21 reviews included in the assessment, 12 reviews considered disease modifying anti-rheumatic drugs (DMARDs), eight considered biologics and one considered glucocorticoids. All reviews required an ORB assessment for at least one eligible trial. The 21 reviews included a total of 172 trials for assessment (Figure 1).

3.1.2 Trial assessed

Among the 172 trials included within the 21 reviews, 17 trials could not be assessed further either because the articles were not in English (n=10) or the trial reports were unobtainable (n=7). Therefore, in our study we assessed 155 trials, 94 on DMARDs, 45 on biologics and 16 on glucocorticoids (Figure 1).

3.2 Assessment results: classification according to ORBIT criteria

Of the 155 assessable trials, 21 trials fully reported the outcome data for all core outcomes but the data in only 10 of these were adequately reported in the review. The ORBIT classifications for all the trials assessed are shown in Table 2. A breakdown of the classification by intervention class (DMARDs, biologics, and glucocosteroids) is provided in web appendix 1 (Table 2).

At the trial level, missing or incomplete reporting of outcome data for each core outcome ranged from 36% (radiological damage 12 out of 33 trials with follow-up greater than 52 weeks) to 56% (physician global 87 out of 155 trials).

For 515 (46%) of the 1118 evaluable outcomes in our study, the set of core outcomes was either partially reported or not reported (A to I classification) (Table 2). For 22% (247 of the 1118 assessable outcomes), at least one core outcome was classified under high suspicion for outcome reporting bias (A, D, E, or G classification), while for 19% (212/1118), it was clear the outcomes were measured and analysed (A, B, C, D classification) but the reporting of the outcomes meant that the data could not be included in a meta-analysis.

3.4 Application to the review of Auranofin for the treatment of rheumatoid arthritis

We now consider one of the reviews in detail [15], which evaluated the evidence for Auranofin for the treatment of rheumatoid arthritis. Of the nine included trials in this review, none fully reported data on all the core outcomes according to OMERACT recommendation. The amount of missing participant data from the original review meta-analysis ranged from 10% (acute phase reactant: ESR) to 66% (Function: HAQ). Table 3 shows the univariate (UFMA, URMA) meta-analysis and multivariate (MFMA, MRMA) meta-analysis results. The within-study correlation between treatment effects estimates for pairs of outcomes was often high and ranged between 0.18 (between swollen joints and acute phase reactant) and 0.91 (between pain and patient global) (web appendix1 (Table 3)).

Comparison of univariate and multivariate fixed effect results

For Auranofin example, high risk ORB classifications were awarded to at least one trial where the MFMA approach provided a treatment estimate that demonstrated a smaller benefit in treatment, which indicates that the UFMA estimates might have been overestimated as a result of ORB.

The UFMA model results were the same as those presented in the original review. When fitting the MFMA model, there was a reduction in standard errors for all outcomes ranging from 3% for APR to 49% for physician global. Of clinical importance, there was a shift in the pooled treatment effect for all of the core outcomes. For example, considering function (HAQ), the pooled mean difference was equal to -0.13 in the UFMA model compared with -0.16 in the MFMA model. The UFMA

estimate showed a benefit for Auranofin (-0.13 change) with a marginal non-significant result (p=0.066), but the MFMA indicates a larger difference between the treatment groups (-0.16 change). As a consequence of the larger difference in treatment effect, coupled with the increased precision, the MFMA estimate was statistically significant (p=0.004). The shift in direction of the estimate when applying MFMA does not suggest ORB as the direction indicates a more positive result. In this example, for this particular outcome, it was unlikely that ORB was present as the studies not reporting on function were all classified as low risk ORB (H-classification).

While there were no changes in statistical significance for any of the other outcomes between the UFMA and MFMA approaches, potentially clinical important differences were found between the treatment effect estimates. There were larger benefit differences in treatment effect favouring Auranofin using the MFMA approach for pain (0.06 difference) but smaller benefit differences in treatment effect for all the other outcomes; the largest difference was for swollen joints where the MFMA approach showed a 84% reduction on benefit (mean difference -0.290) for Auranofin when compared to the estimate from the URMA model (mean difference -0.047). The smallest change in the pooled treatment effect between the two models was for acute phase reactants (APR), where there was almost complete data. For APR the MRMA approach showed a 4% reduction on benefit (mean difference -9.040) for Auranofin when compared to the estimate from the URMA model (mean difference -8.723).

Comparison of univariate and multivariate random effects results

When we fitted the MRMA model, the I^2_R value was 83% indicating that across all outcomes, the total variation in the meta-analysis is mainly due to between-study heterogeneity; this suggests that the random-effects model might be the appropriate model to fit to these data. For the MRMA model we observed smaller standard errors and hence improved statistical precision for some, but not all, of the outcomes when compared to the URMA model. Some outcomes had lower precision of their summary effect in the MRMA because the between study variance estimates becoming larger than those from URMA. There was no change in statistical significance at the 5% level for any of the outcomes when comparing URMA with MRMA; however, as seen in the fixed effect results, the outcome function (HAQ) was closer to statistical significance (p=0.054) favouring Auranofin when applying the MRMA rather than URMA model. Also there remain large differences between the

summary treatment effect estimates when comparing URMA and MRMA (largest increase is equal to 176% for swollen joints). In contrast to the fixed effect models, neither of the global measurements (physician and patient) reaches statistical significance for either of URMA or MRMA, the likely result of the summary treatment effects having larger standard errors due to larger estimates of between-study heterogeneity. When we applied the MRMA, the summary treatment estimates moved towards the null for tender joints and pain where high suspicion ORB was suspected. The shift in treatment effect estimates was in the opposite direction for the global measurements (physician and patient), function and acute phase reactant. The impact of ORB was likely to be minimal for these outcomes.

4. Discussion

This is the first study to consider an assessment of outcome reporting bias against a well-established core set of outcomes. The OMERACT core outcome set (sometimes referred to as ILAR [International League of Associations for Rheumatology] core set of outcomes) was ratified for use in clinical trials of rheumatoid arthritis but is also endorsed by the Cochrane Musculoskeletal Group [19]. While the uptake of the measurement of the core outcome set for rheumatoid arthritis trials has been shown to be increasing [7], the reporting of the outcomes for many of these remains insufficient, meaning that many meta-analyses are unable to include data from all relevant studies. Similar to the [4] study that looked at all review outcomes in a cohort of Cochrane Fibrosis reviews, all the reviews considered in this study included at least one study which contained missing data on at least one core outcome. Across all core outcomes, there were 212 items of study data missing from meta-analyses for outcomes that were clearly measured and analysed but either not reported or reported inappropriately (A-D classifications), and a further 191 items of study data were clearly measured or likely measured but not reported because of non-significant results (E and G classifications). It is important that trialists follow CONSORT 2010 [20] (Consolidated Standards of Reporting Trials) guidance for reporting trials findings. Adherence to the CONSORT statement would ensure that all outcomes are reported in full and all pre-specified outcomes are defined and reported.

Many outcomes were not mentioned in trial reports meaning that clinical judgement was needed to decide whether the outcome of interest was likely to have been measured for a particular trial. A

limitation of this study was that we did not contact trialists to determine whether outcomes were measured if they were not mentioned although any uncertainties in classifications were confirmed by contact with review authors. Our decision not to contact trialists in this study was a pragmatic one. The most recent trial published for inclusion in this study was published over five years ago (median publication date 1999), meaning that there would be obvious difficulties in locating the majority of trialists. Nevertheless, the reliability of systematic reviews can be improved if more attention is paid to outcome data missing from the source trial reports. If data are missing, reviewers should be encouraged to at least attempt to contact the trialists or study sponsors to confirm whether the outcome was measured and analysed and, if so, obtain the results and update the review meta-analysis accordingly with the newly obtained data. Reviewers should also be encouraged to complete the Cochrane risk of bias tool. A new version the Cochrane risk of bias tool which includes a section on '*bias in selection of reported result'* which is informed by the ORBIT study and is set to be launched in 2014 [21]. If obtaining outcome data is not feasible or successful then rather than do nothing, review authors are encouraged to apply a sensitivity analysis to assess the impact of outcome reporting bias on an individual review.

The multivariate meta-analysis approach offers one such sensitivity analysis to reduce the potential impact of outcome reporting bias when there is missing trial data for many review outcomes. Our recommendation to reviewers would be to use the multivariate meta-analysis approach if, as in the Auranofin example, reasonable correlation estimates between outcomes are available. If estimates from IPD are not available then, one could also consider clinical or biological reason to inform the correlation, or consider sensitivity analyses over a range of sensible values [9]. For example, in the rheumatoid arthritis case study, we would hypothesise that the relationship between the number of tender joints has a positive correlation with pain (more tender joints would imply more pain), although the strength of the correlation may be more difficult to judge. There is a need for a central repository for individual participant data from trials [22]. This would not only greatly reduce the amount of missing data from reviews and reduce the possibility of ORB but also provide review authors with the means to obtain reliable within-study estimates of correlations (also for use in studies where this information is not available) should an analytical technique for adjusting a meta-analysis be required. If one is not confident about the correlations between outcomes, then other

univariate sensitivity approaches under an assumed model for selective reporting could be applied to each outcome, for example the method performed by Copas et al. [1].

The Auranofin example was used to demonstrate the MVMA approach for two reasons. Firstly high-risk ORB classifications were awarded to at least one trial for tender joints count, pain, physician global and acute phase reactant. Secondly, we observed potentially important statistical and clinical differences when comparing the treatment estimates between MVMA with univariate meta-analysis. In some instances the differences between the univariate and MVMA results is minimal; this may result when there is not much missing outcome data from the reviews or ORB suspicion is low [10]. The results to the application of the MVMA approach to all other rheumatoid arthritis reviews where an ORB assessment was considered in this study is available from the corresponding author on request. When applying the MVMA method to the Auranofin review, we choose to explore what happens when we specify both fixed and random effect models as we wanted to demonstrate the MVMA method using both model specifications. However, in practice, the choice of a fixed effect or a random effect model should be specified in the review protocol. Application of the multivariate meta-analysis approach to the Auranofin example has demonstrated a change in the treatment estimates between univariate and multivariate meta-analysis. High risk ORB classifications were awarded to at least one trial for tender joints count, pain, physician global and acute phase reactant. When we applied a multivariate meta-analysis (fixed or random), the pooled effect estimates moved towards the null for tender joints, swollen joints (fixed only), pain (random only), physician global (fixed only), patient global (fixed only) and acute phase reactant (fixed only). This shift in direction of the treatment effect is suggestive that ORB maybe an issue for these outcomes. When the shift in treatment effects estimate is in the opposite direction then the impact of ORB is likely to be minimal.

A previous study, which compared univariate meta-analysis with multivariate meta-analysis, led the authors to conclude that the choice between univariate and multivariate meta-analysis had unclear or limited practical importance [13]. Though this may often be the case, our evaluation suggests that multivariate meta-analysis approach may be especially relevant to adjust for outcome reporting bias when there is missing trial data for many review outcomes. Therefore our recommendation to

reviewers would be to use the multivariate meta-analysis approach if, as in the Auranofin example, reasonable correlation estimates between outcomes are available and ORB is a genuine concern.

References

[1] Copas J, Dwan K, Kirkham JJ and Williamson PR. A model-based correction for outcome reporting bias in meta-analysis. Biostatistics 2014; 15(2):370-83. DOI:10.1093/biostatistics/kxt046.

[2] Dwan K, Gamble C, Williamson PR and Kirkham JJ. Systematic review of the empirical evidence of study publication bias and outcome reporting bias - An updated review. PLoS ONE 2013; 8(7): e66844. DOI:10.1371/journal.pone.0066844.

[3] Kirkham JJ, Dwan K, Altman DG, Gamble C, Dodd S, Smyth R and Williamson PR. The impact of outcome reporting bias in randomised controlled trials on a cohort of systematic reviews. BMJ 2010; 340:c365. DOI:10.1136/bmj.c365.

[4] Dwan K, Kirkham JJ, Williamson PR and Gamble C. Selective reporting of outcomes in randomised controlled trials in systematic reviews of cystic fibrosis. BMJ Open 2013; 3:e0027092013. DOI:10.1136/bmjopen-2013-002709.

[5] Williamson PR, Altman DG, Blazeby JM, Clarke M and Gargon E. Driving up the quality and relevance of research through the use of agreed core outcomes. J Health Serv Res Policy 2012; 17(1). DOI:10.1258/jhsrp.2011.011131.

[6] Boers M, Tugwell P, Felson DT, van Riel PL, Kirwan JR, Edmonds JP, Smolen JS, Khaltaev N, and Muirden KD. World Health Organization and International League of Associations for Rheumatology core endpoints for symptom modifying antirheumatic drugs in rheumatoid arthritis clinical trials. J Rheumatol Suppl. 1994; 41:86-9.

[7] Kirkham JJ, Boers M, Tugwell P, Clarke M and Williamson PR. Outcome measures in rheumatoid arthritis randomised trials over the last 50 years. Trials 2013; 14:324. DOI: 10.1186/1745-6215-14-324.

[8] Van der Heijde DM, Van't Hof M, Van Riel PL and Van de Putte LB. Development of a disease activity score based on judgment in clinical practice by rheumatologists. J Rheumatol.1993; 20(3):579–581.

[9] Felson DT, Anderson JJ, Boers M, Bombardier C, Chernoff M, Fried B, Furst D, Goldsmith C, Kieszak S, Lightfoot R, Paulus H, Tugwell P, Weinblatt M, Widmark R, Williams JH and Wolfe F. The American college of rheumatology preliminary core set of disease activity measures for rheumatoid arthritis clinical trials. Arthritis Rheum.1993; 36(6):729–740. DOI:10.1002/art.1780360601.

[10] Jackson D, White I and Riley RD. Quantifying the impact of between-study heterogeneity in multivariate meta-analyses. *Statistics in Medicine* 2012; 31(29):3805–3820. DOI:10.1002/sim.5453.

[11] Kirkham JJ, Riley RD and Williamson PR. A multivariate meta-analysis approach for reducing the impact of outcome reporting bias in systematic reviews. Statistics in Medicine 2012; 31(20): 2179-2195. DOI:10.1002/sim.5356.

[12] Riley RD. Multivariate meta-analysis: the effect of ignoring within-study correlation. Journal of the Royal Statistical Society: Series A 2009; 172(4):789–811. DOI: 10.1111/j.1467-985X.2008.00593.x.

[13] Trikalinos TA, Hoaglin DC and Schmid CH. An empirical comparison of univariate and multivariate meta-analyses for categorical outcomes. Statistics in Medicine 2013; 33(9):1441-1459 DOI: 10.1002/sim.6044.

[14] Dwan K, Gamble C, Kolamunnage-Dona R, Mohammed S, Powell C and Williamson PR. Assessing the potential for outcome reporting bias in a review: a tutorial. Trials 2010; 11:52. DOI:10.1186/1745-6215-11-52.

[15] Suarez-Almazor ME, Spooner C, Belseck E, Shea B (2009). Auranofin versus placebo in rheumatoid arthritis (Review). The Cochrane Library, Issue 1, CD002048.

[16] Riley RD1, Higgins JP, Deeks JJ. Interpretation of random effects meta-analyses. BMJ 2011 Feb 10; 342:d549. DOI: 10.1136/bmj.d549.

[17] Higgins JPT, Thompson SG, and Spiegelhalter DJ. A re-evaluation of random-effects meta-analysis. J R Stat Soc Ser A Stat Soc. Jan 2009; 172(1): 137–159. DOI: 10.1111/j.1467-985X.2008.00552.x. PMCID: PMC2667312.

[18] White I. Multivariate random-effects meta-analysis. The Stata Journal (2009), 9, Number 1, pp. 40-56.

[19] Tugwell P, Boers M, Brooks P, Simon L, Strande V and Idzerda L. OMERACT: an international initiative to improve outcome measurement in rheumatology. Trials 2007; 8:38. DOI: 10.1186/1745-6215-8-38.

[20] Schulz KF, Altman DG and Moher D. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. BMJ 2010; 340:c332. DOI:10.1136/bmj.c332.

[21] Turner L, Boutron I, Hróbjartsson A, Altman DG and Moher D. The evolution of assessing bias in Cochrane systematic reviews of interventions: celebrating methodological contributions of the Cochrane Collaboration. Systematic Reviews 2013; 2:79. DOI: 10.1186/2046-4053-2-79.

[22] Tudur-Smith C, Dwan K, Altman DG, Clarke M, Riley R, Williamson PR. Sharing individual participant data from clinical trials: an opinion survey regarding the establishment of a central repository. PloS ONE 2014; 9 (5): e97886.

Acknowledgements:

We would like to thank Dr George Wells for providing the correlations between the core outcomes which were obtained from individual patient data.

Funding:

This work forms part of the first author's PhD funded by Liverpool Reviews and Implementation Group (LRiG), Department of Biostatistics, University of Liverpool and Institute of Translational Medicine, University of Liverpool. RDR and JJK are supported by funding from a Multivariate meta-analysis grant from the MRC Methodology Research Programme (grant reference number: MR/J013595/1).

Table.1 The Outcome Reporting Bias in Trials (ORBIT) study classification for missing or incomplete reporting in reports of randomized controlled trials

	Description	Level of reporting	Risk of bias
Clear that the outcome was measured and analysed			
A	Trial report states that outcome was analysed but only reports that result was not significant (typically stating p-value>0.05).	Partial	High Risk
B	Trial report states that outcome was analysed but only reports that result was significant (typically stating p-value<0.05).	Partial	No Risk
C	Trial report states that outcome was analysed but insufficient data were presented for the trial to be included in meta-analysis or to be considered to be fully tabulated.	Partial	Low Risk
D	Trial report states that outcome was analysed but no results reported.	None	High Risk
Clear that the outcome was measured			
E	Clear that the outcome was measured. Judgment says outcome ***likely*** to have been analysed but not reported because of non-significant results.	None	High Risk
F	Clear that the outcome was measured. Judgment says outcome ***unlikely*** to have been analysed.	None	Low Risk
Unclear whether the outcome was measured			
G	Not mentioned but clinical judgment says likely to have been measured and analysed but not reported on the basis of non-significant results.	None	High Risk
H	Not mentioned but clinical judgment says unlikely to have been measured at all.	None	Low Risk
Clear that the outcome was not measured			
I	Clear that the outcome was not measured.	NA	No Risk

Risk of bias arising from the lack of inclusion of non-significant results when a trial was excluded from a meta-analysis or non-fully reported in a review because the data were unavailable.

Table 2: Clinical trials assessed for outcome reporting bias (n=155 trials)

Classification	Rheumatoid Arthritis Core Set of Outcomes									TOTAL (%)[d]
	Tender Joints	Swollen Joints	Pain	Patient Global	Physician Global	Function	APR	RD		
								< 52 weeks	≥ 52 weeks	
A	3	4	1	3	2	1	4	2	2	20 (1.8)
B	1	1	0	0	1	0	0	0	0	3 (0.3)
C	21	18	17	20	18	23	34	1	2	153 (13.7)
D	2	1	6	6	7	7	5	2	2	36 (3.2)
E	17	15	17	20	19	7	13	2	0	108 (9.7)
F	9	9	8	7	6	7	8	0	0	54 (4.8)
G	4	11	20	15	18	6	3	5	6	83 (7.4)
H	1	3	11	11	16	13	3	81	0	58 (5.2)
I	0	0	0	0	0	0	0	4	0	0 (0.0)
Fully reported	97	93	75	73	68	91	85	25	21	603 (54)
Total	155	155	155	155	155	155	155	122	33	1118
TOTAL Missing Data (A-I) (%)[a]	58 (37.4)	62 (40.0)	80 (51.6)	82 (52.9)	87 (56.1)	64 (41.3)	70 (45.2)	N/A[b]	12 (36.4)[c]	

APR: Acute Phase Reactant; RD: Radiological Damage.

[a] The denominator used is the total number of trials where an assessment is possible (155).

[b] Not applicable: OMERACT recommends outcome only applicable if follow-up > 52 weeks.

[c] The denominator used is the total number of trials where an assessment is possible and the follow-up is greater than 52 weeks (33) .

[d] The denominator used is 1118. That is the total number of data points if all 155 trials reported on all seven core outcomes (TJC, SJC, Pain, Pat. Global, Phy.Global, Function, APR) plus the 33 trials that should have also measured and reported on RD due to a follow-up greater than 52 weeks (i.e. (155*7) + (33*1) = 1118). The numerator also excludes the assessment of RD for trials less than 52 weeks.

Table 3: Meta-analysis results for the Auranofin dataset

	Univariate meta-analysis				Multivariate meta-analysis			
	Mean Diff. †	**S.E.** ††	**p-value**	**95% C.I.** †††	**Mean Diff.**	**S.E.** †	**p-value**	**95% C.I.** ††
	UFMA				**MFMA**			
Y1	-3.759	0.666	<0.001***	-5.064; -2.453	-3.225	0.561	<0.001***	-4.324; -2.126
Y2	-0.290	0.608	0.634	-1.482; 0.902	-0.047	0.559	0.933	-1.143; 1.049
Y3	-4.681	0.974	<0.001***	-6.589; -2.772	-4.998	0.565	<0.001***	-6.105; -3.891
Y4	-0.368	0.081	<0.001***	-0.527; -0.210	-0.275	0.062	<0.001***	-0.395; -0.154
Y5	-0.409	0.120	0.001**	-0.645; -0.173	-0.244	0.061	<0.001***	-0.363; -0.125
Y6	-0.130	0.071	0.066	-0.268; 0.008	-0.160	0.055	0.004**	-0.268; -0.052
Y7	-9.040	1.592	<0.001***	-12.159; -5.920	-8.723	1.540	<0.001***	-11.741; -5.705
	URMA				**MRMA**			
Y1	-3.822	0.921	<0.001***	-5.627; -2.016	-3.621	0.967	<0.001***	-5.516; -1.727
Y2	0.148	1.181	0.900	-2.168; 2.463	0.408	1.166	0.726	-1.877; 2.693
Y3	-4.681	0.974	<0.001***	-6.589; -2.772	-4.438	1.065	<0.001***	-6.525; -2.351
Y4	-0.392	0.207	0.058	-0.797; 0.014	-0.394	0.253	0.120	-0.891; 0.102
Y5	-0.409	0.120	0.001**	-0.645; -0.173	-0.429	0.279	0.124	-0.976; 0.118
Y6	-0.130	0.071	0.066	-0.268; 0.008	-0.208	0.108	0.054	-0.420; 0.003
Y7	-9.788	3.252	0.003**	-16.162; -3.414	-10.259	3.221	0.001***	-16.572; -3.946

Y1: Tender Joints Count; Y2: Swollen Joints; Y3: Pain; Y4: Physician Global; Y5: Patient Global; Y6: Function; Y7: Acute Phase Reactant.

† Mean Diff.: Mean Difference

†† S.E.: Standard Error of summary treatment effect estimate.

††† C.I.: Confidence Interval

*p-value<0.05; **p-value<0.005; ***p-value<0.001

Figure.1: Flow diagram of reviews eligibility and assessment of randomised controlled trials within reviews